

LA-UR-21-30744

Approved for public release; distribution is unlimited.

Title: Robustify Your Links! For better stewardship of references to web resources in digital scholarship

Author(s): Klein, Martin

Intended for: LANL RL OA week presentation

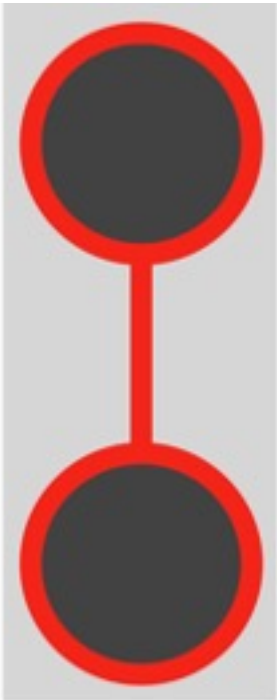
Issued: 2021-10-28

Disclaimer:

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Robustify Your Links!

For better stewardship
of references to web resources
in digital scholarship



Martin Klein

mklein@lanl.gov

[@mart1nkle1n](https://twitter.com/mart1nkle1n)

<http://orcid.org/0000-0003-0130-2097>

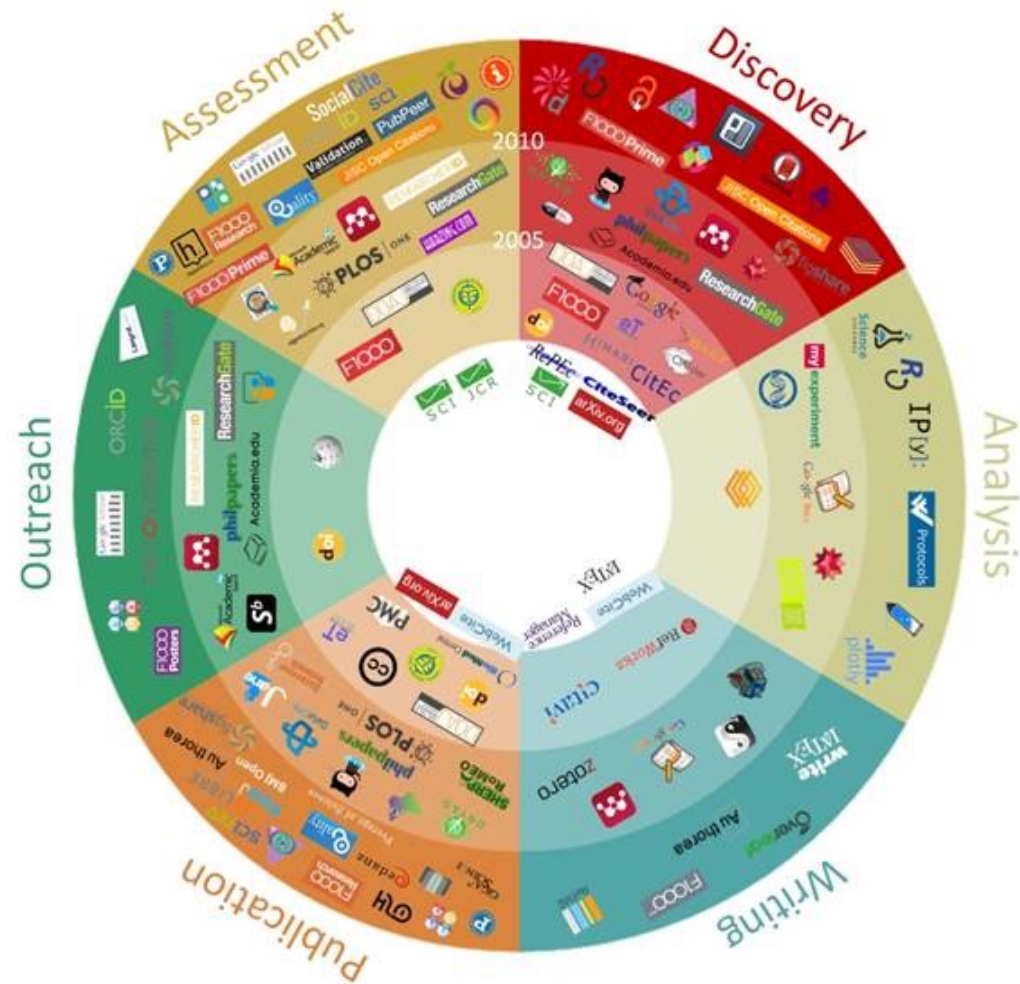
Research Library
Los Alamos National Laboratory

```
<a href="http://www.whitehouse.gov"
data-versionurl="https://perma.cc/9265-T4NB"
data-versiondate="2015-11-18">
The White House</a>
```

Web-based scholarly communication



101 Innovations in scholarly communication



Jeroen Bosman and Bianca Kramer

<https://101innovations.wordpress.com/>

Scholarly communication questions

- How does this affect our references in scholarly articles?
 - Do we increasingly reference resources on the web?
 - What sort of web resources do we reference?

“Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot”

<https://doi.org/10.1371/journal.pone.0115253>

“Scholarly Context Adrift: Three out of Four URI References Lead to Changed Content”

<https://doi.org/10.1371/journal.pone.0167475>

Scholarly communication concerns

- What happens to such references over time?
 - Do they remain stable?
 - Do they break?
 - Does their content change?

Stable reference

Reference URI:

http://ifa.hawaii.edu/~cowie/k_table.html

Referenced in:

<http://arxiv.org/abs/astro-ph/9707064>

published on July 4th 1997

The screenshot shows the Internet Archive interface for the URL http://www.ifa.hawaii.edu/~cowie/k_table.html. The page title is "H+K ordered Table". The text explains that magnitudes are measured in 3 arcsecond diameter apertures and corrected to 6 arcsecond diameters. It lists the zero points for the bands: Johnson K, J, V(6020) and B(4500), Kron Cousins I(8140) and AB for the 3000. The H+K catalog slightly clips the HDF. Spectroscopic redshifts $z(sp)$ are all from the Hawaii and Caltech samples except for objects with $z \geq 2.4$ where the values are from Steidel et al. (1996). $z(6)$ is an SED estimated redshift from the 6 colors and for this column $z=3$ means that there are two distinct objects corresponding to the K image and no SED redshift can be obtained using the six color estimator. A finding chart for the objects on the HDF814W image is given in `hdf_k_chart`. *** means an object was not detected in that band.

Objects 180 through 183 expand the K<22 table to include all B<25 objects.

All additional Lick redshifts added October 3rd 1996. (objects 47,100,110,118,136,151)

R.S. Ellis type added from Van den Bergh et al. A.J. paper for classified objects with $K < 21.5$ or $I < 23.5$. -1 indicates object was not located in table. (0=ell,1=spiral,2=irr)

Redshifts and spectra for objects 21,40,45,46,59,70,76,81,84,93,95,109,131 added March 18th 1997 based on Hawaii March LRIS run.

#	RA(2000)	DEC(2000)	H+K	J	8140	6020	4500	3000	z(sp)	z(6)	RSetype
0	12 36 49.33	62 13 47.81	16.0	16.6	17.7	19.0	20.0	22.7	0.089	0.131	1
1	12 36 54.65	62 13 29.07	16.8	17.4	18.8	20.2	21.4	25.5	-2.000	-2.000	-2
2	12 36 56.26	62 12 42.35	17.3	17.7	18.6	19.6	20.0	22.2	-2.000	-2.000	-2
3	12 36 56.56	62 12 46.84	17.5	18.6	19.9	21.8	23.6	25.6	0.517	0.548	1
4	12 36 48.00	62 13 10.00	17.6	18.6	19.9	21.6	23.4	26.6	0.475	0.489	0
5	12 36 43.75	62 11 43.91	17.7	18.7	20.3	22.5	24.6	26.1	0.764	0.807	0
6	12 36 51.00	61 13 21.56	17.9	18.7	19.4	20.4	21.0	22.0	0.199	0.091	1
7	12 36 53.80	62 12 55.17	17.9	19.1	20.3	22.0	22.9	23.7	0.642	0.648	1
8	12 36 46.14	62 11 43.16	18.0	19.1	20.8	22.4	23.0	23.5	1.012	0.827	1
9	12 36 46.26	62 14 5.54	18.0	19.0	20.7	22.6	23.8	25.9	0.960	0.787	0
10	12 36 51.70	62 13 54.79	18.0	19.2	20.5	22.0	23.1	24.1	0.557	0.648	2
11	12 36 42.84	62 12 17.24	18.2	19.1	20.1	21.3	22.3	22.8	0.454	0.489	1
12	12 36 43.08	62 12 43.13	18.2	19.4	21.1	23.4	25.3	26.4	0.847	0.886	0
13	12 36 50.18	62 12 46.86	18.2	19.4	20.8	23.0	24.7	26.5	0.678	0.687	0
14	12 36 41.85	62 12 6.28	18.2	19.2	20.4	21.7	22.6	23.2	0.432	0.508	1
15	12 36 50.15	62 12 40.83	18.3	19.2	20.1	21.3	22.2	22.7	0.474	0.369	0
16	12 36 55.36	62 13 12.33	18.3	19.6	21.6	24.0	25.9	****	0.000	1.006	1
17	12 36 49.64	62 13 14.04	18.4	19.7	21.0	22.4	23.7	25.1	0.475	0.489	1
18	12 37 0.47	62 12 35.94	18.4	19.5	20.8	22.5	24.2	26.4	0.562	0.528	0
19	12 36 51.64	62 12 21.34	18.6	19.6	20.9	22.3	23.4	24.8	0.299	0.290	0

June 7th 1997

The screenshot shows the current version of the ifa.hawaii.edu/~cowie/k_table.html page. The title is "H+K ordered Table". The text explains that magnitudes are measured in 3 arcsecond diameter apertures and corrected to 6 arcsecond diameters with an average offset for that color. Positions are measured on the H+K image. The zero points for the bands are Johnson K, J, V(6020) and B(4500), Kron Cousins I(8140) and AB for the 3000. The H+K catalog slightly clips the HDF. Spectroscopic redshifts $z(sp)$ are all from the Hawaii and Caltech samples except for objects with $z \geq 2.4$ where the values are from Steidel et al. (1996). $z(6)$ is an SED estimated redshift from the 6 colors and for this column $z=3$ means that there are two distinct objects corresponding to the K image and no SED redshift can be obtained using the six color estimator. A finding chart for the objects on the HDF814W image is given in `hdf_k_chart`. *** means an object was not detected in that band.

Objects 180 through 183 expand the K<22 table to include all B<25 objects.

All additional Lick redshifts added October 3rd 1996. (objects 47,100,110,118,136,151)

R.S. Ellis type added from Van den Bergh et al. A.J. paper for classified objects with $K < 21.5$ or $I < 23.5$. -1 indicates object was not located in table. (0=ell,1=spiral,2=irr)

Redshifts and spectra for objects 21,40,45,46,59,70,76,81,84,93,95,109,131 added March 18th 1997 based on Hawaii March LRIS run.

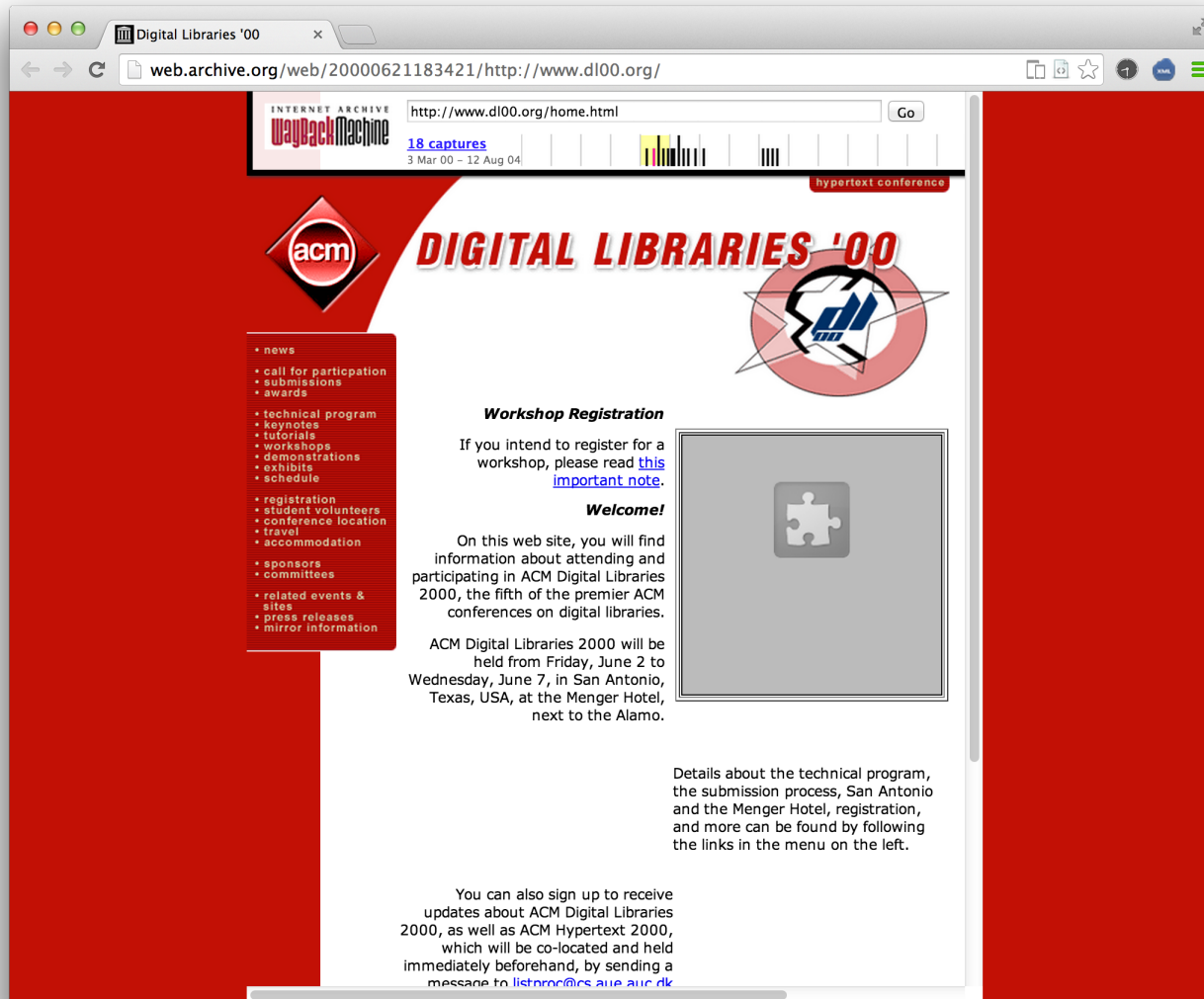
#	RA(2000)	DEC(2000)	H+K	J	8140	6020	4500	3000	z(sp)	z(6)	RSetype
0	12 36 49.33	62 13 47.81	16.0	16.6	17.7	19.0	20.0	22.7	0.089	0.131	1
1	12 36 54.65	62 13 29.07	16.8	17.4	18.8	20.2	21.4	25.5	-2.000	-2.000	-2
2	12 36 56.26	62 12 42.35	17.3	17.7	18.6	19.6	20.0	22.2	-2.000	-2.000	-2
3	12 36 56.56	62 12 46.84	17.5	18.6	19.9	21.8	23.6	25.6	0.517	0.548	1
4	12 36 48.00	62 13 10.00	17.6	18.6	19.9	21.6	23.4	26.6	0.475	0.489	0
5	12 36 43.75	62 11 43.91	17.7	18.7	20.3	22.5	24.6	26.1	0.764	0.807	0
6	12 36 51.00	61 13 21.56	17.9	18.7	19.4	20.4	21.0	22.0	0.199	0.091	1
7	12 36 53.80	62 12 55.17	17.9	19.1	20.3	22.0	22.9	23.7	0.642	0.648	1
8	12 36 46.14	62 11 43.16	18.0	19.1	20.8	22.4	23.0	23.5	1.012	0.827	1
9	12 36 46.26	62 14 5.54	18.0	19.0	20.7	22.6	23.8	25.9	0.960	0.787	0
10	12 36 51.70	62 13 54.79	18.0	19.2	20.5	22.0	23.1	24.1	0.557	0.648	2
11	12 36 42.84	62 12 17.24	18.2	19.1	20.1	21.3	22.3	22.8	0.454	0.489	1
12	12 36 43.08	62 12 43.13	18.2	19.4	21.1	23.4	25.3	26.4	0.847	0.886	0
13	12 36 50.18	62 12 46.86	18.2	19.4	20.8	23.0	24.7	26.5	0.678	0.687	0
14	12 36 41.85	62 12 6.28	18.2	19.2	20.4	21.7	22.6	23.2	0.432	0.508	1
15	12 36 50.15	62 12 40.83	18.3	19.2	20.1	21.3	22.2	22.7	0.474	0.369	0
16	12 36 55.36	62 13 12.33	18.3	19.6	21.6	24.0	25.9	****	0.000	1.006	1
17	12 36 49.64	62 13 14.04	18.4	19.7	21.0	22.4	23.7	25.1	0.475	0.489	1
18	12 37 0.47	62 12 35.94	18.4	19.5	20.8	22.5	24.2	26.4	0.562	0.528	0
19	12 36 51.64	62 12 21.34	18.6	19.6	20.9	22.3	23.4	24.8	0.299	0.290	0

today

Link Rot



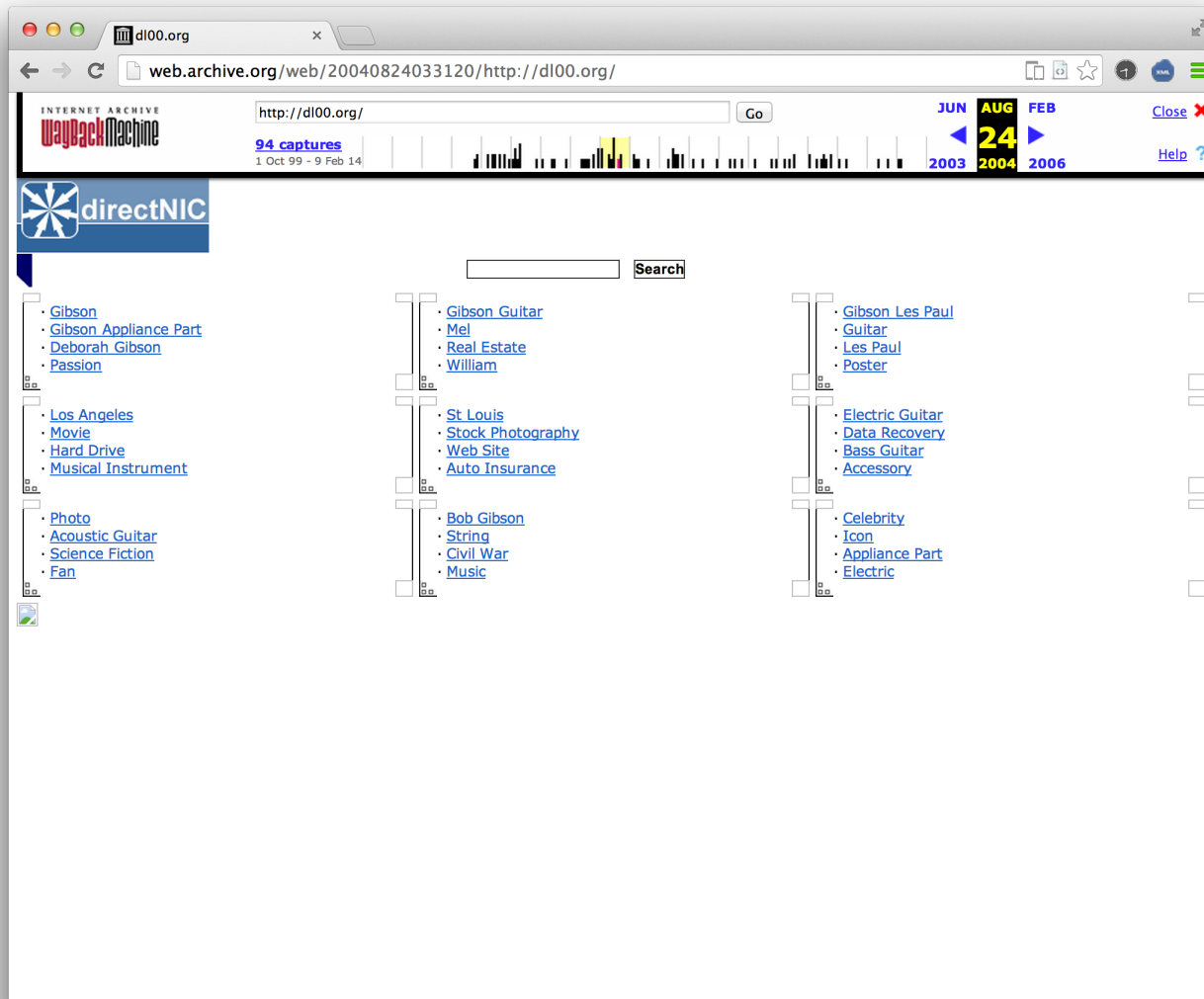
Content Drift



http://dl00.org

2000

Content Drift

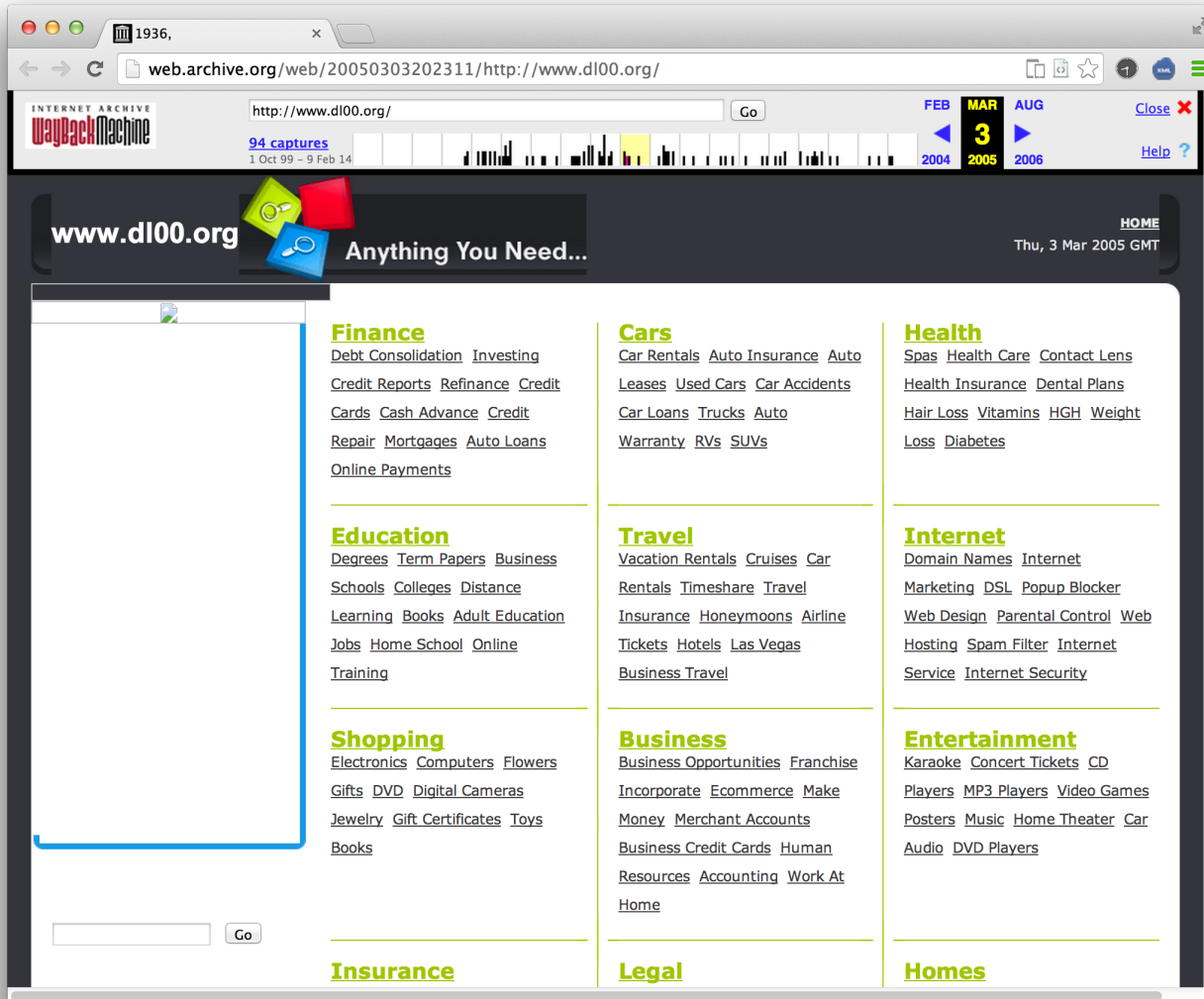


<http://dl00.org>

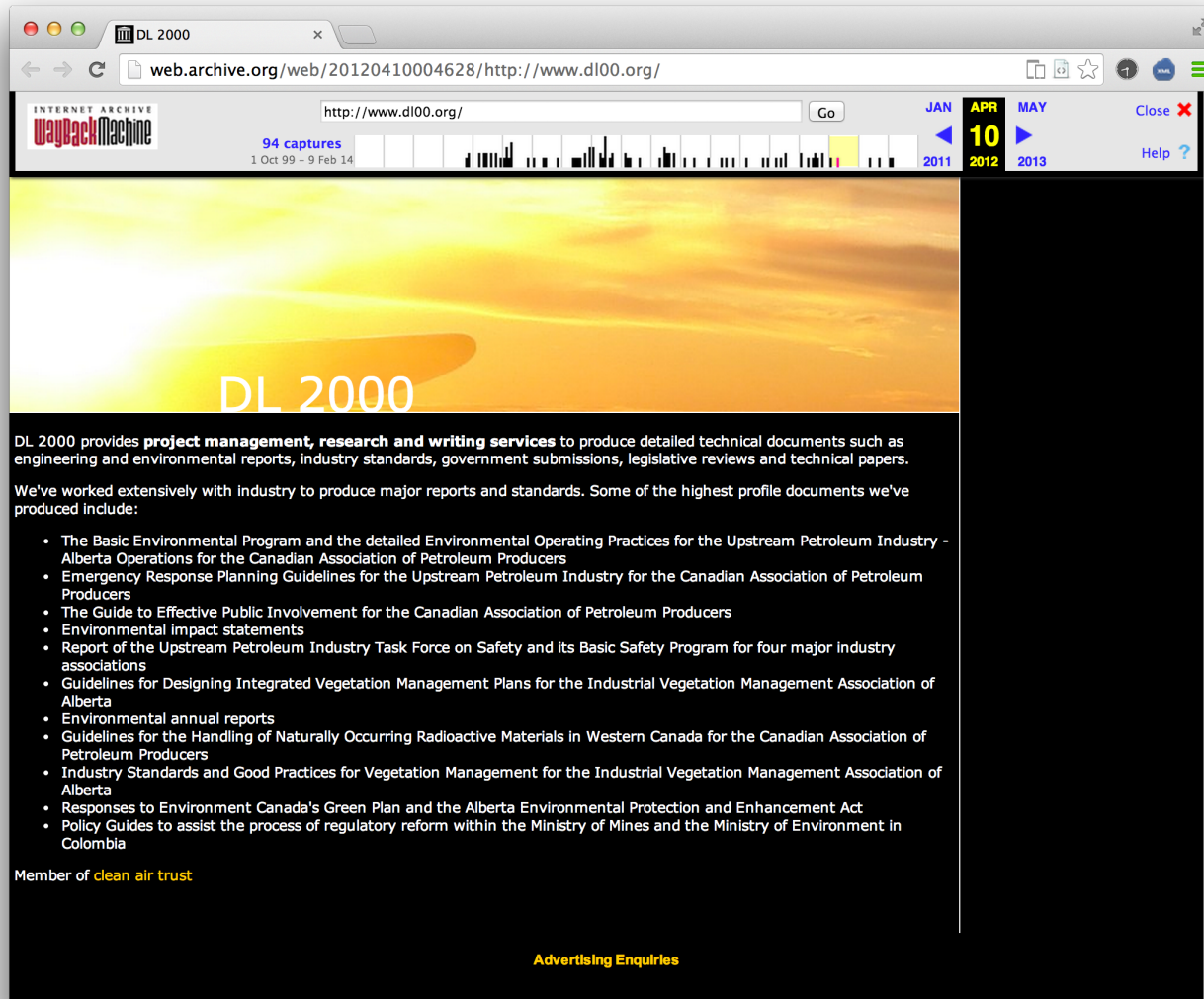
2004

Content Drift

<http://dl00.org>
2005



Content Drift



<http://dl00.org>
2008

Reference Rot

- Link Rot: Link stops working e.g., HTTP 404
- Content Drift: Linked content changes over time
- Reference Rot = Link Rot + Content Drift



<http://hiberlink.org/>

Reference Rot in scholarly communication

- Web resources referenced in scholarly articles:
 - Are subject to reference rot just like any other web resource.
 - Are not necessarily under the custodianship of parties that care about long term integrity and access.
 - Do not necessarily have the same sense of fixity that, for example, journal articles have.

Who cares?

The New York Times Cares

SIDEBAR

In Supreme Court Opinions, Web Links to Nowhere

By ADAM LIPTAK

Published: September 23, 2013

WASHINGTON — Supreme Court opinions have come down with a bad case of link rot. According to [a new study](#), 49 percent of the hyperlinks in Supreme Court decisions no longer work.

[Enlarge This Image](#)



Stephan Savola/Associated Press
Justice Samuel A. Alito Jr.

This can sometimes be amusing. A link in [a 2011 Supreme Court opinion](#) about violent video games by Justice Samuel A. Alito Jr. now leads to [a mischievous error message](#).

“Aren’t you glad you didn’t cite to this Web page?” it asks. “If you had, like Justice Alito did, the original content would have long since disappeared and someone else might have come along and purchased the domain in order to make a comment about the transience of linked information in the Internet age.”

The prankster has a point. The modern Supreme Court opinion is increasingly built on sand.

Hyperlinks are a huge and welcome convenience, of course, said [Jonathan Zittrain](#), who teaches law and computer science at Harvard and who prepared the study with [Kendra Albert](#), a law student there. “Things are readily accessible,” he said, “until they aren’t.”

FACEBOOK

TWITTER

GOOGLE+

SAVE

E-MAIL

SHARE

PRINT

REPRINTS

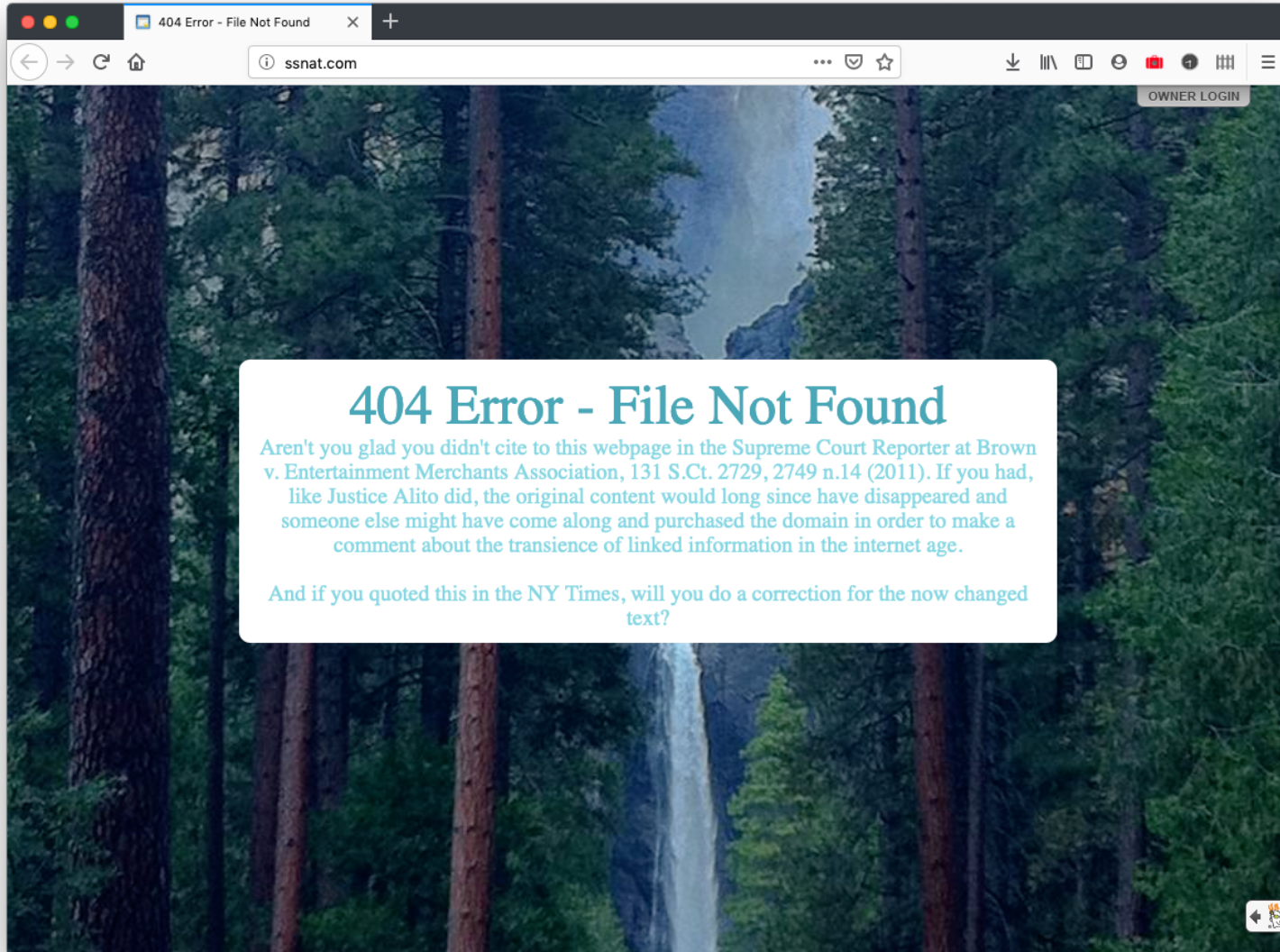
BLACK NATIVITY
NOVEMBER 27
WATCH TRAILER

Links in Supreme Court decisions:

- Link rot: 29%
- Reference rot: 49.9%

<http://www.nytimes.com/2013/09/24/us/politics/in-supreme-court-opinions-clicks-that-lead-nowhere.html>

The Supreme Court should care

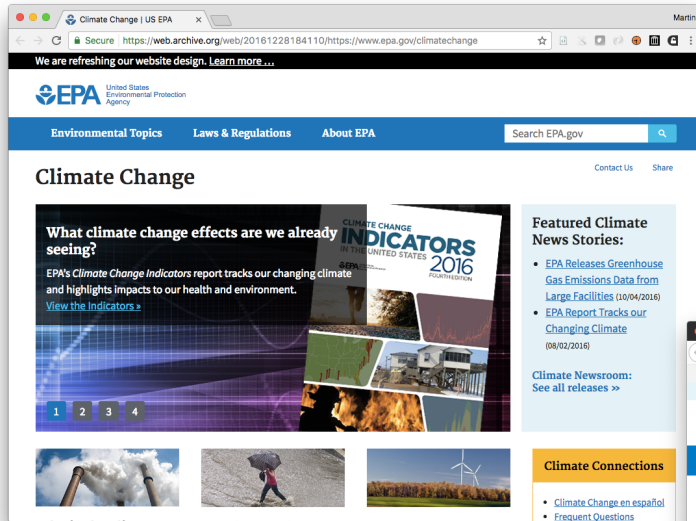


<http://ssnat.com/>

Everybody should care

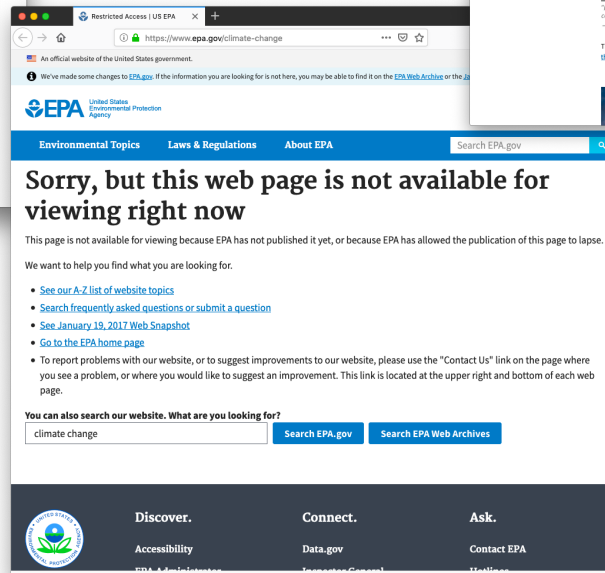
today

12/2016

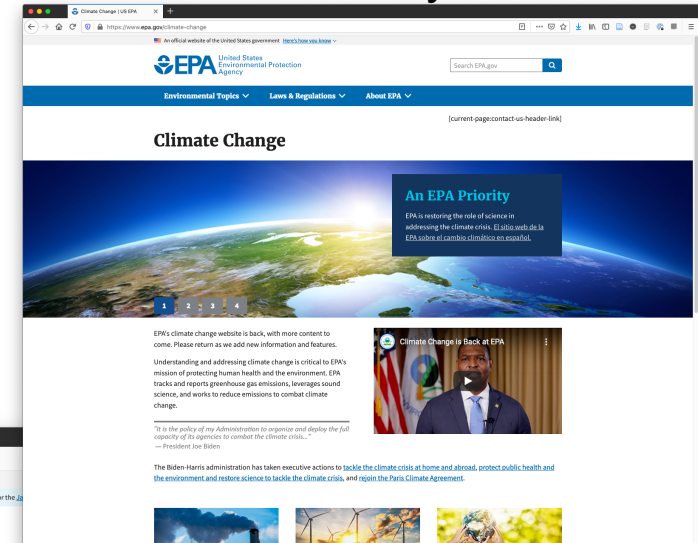


<https://web.archive.org/web/20161228184110/https://www.epa.gov/climatechange>

10/2020



<http://web.archive.org/web/20201027183049/https://www.epa.gov/climate-change>



<https://www.epa.gov/climatechange>

Authors should care

References

D-Lib Magazine
September 2004

Volume 10 Number 9
ISSN 1082-9873

Rethinking Scholarly Communication

Building the System that Scholars Deserve

Herbert Van de Sompel
Los Alamos National Laboratory, Research Library
<herbertv@lanl.gov>

Sandy Payette
Cornell University, Computing and Information Science
<payette@cs.cornell.edu>

John Erickson
Hewlett-Packard Laboratories, Digital Media Systems Lab
<john.erickson@hp.com>

Carl Lagoze
Cornell University, Computing and Information Science
<lagoze@cs.cornell.edu>

Simeon Warner
Cornell University, Computing and Information Science
<simeon@cs.cornell.edu>

<https://doi.org/10.1045/september2004-vandesompel>

!Exist

!Exist

!Exist

!Exist

!Exist

Exist

Atkins, D. et al.. 2003. National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, *Revolutionizing Science and Engineering through Cyber-infrastructure*, <http://www.communitytechnology.org/nsf_ci_report/>.

Brody, T., Kampa, S., Harnad, S., Carr, L. and Hitchcock, S. 2003. Digitometric Services for Open Archives Environments. In *Proceedings of European Conference on Digital Libraries 2003*, pages pp. 207-220, Trondheim, Norway. <<http://eprints.ecs.soton.ac.uk/archive/00007503/>>.

Frey, J., De Roure, D. and Carr, L. 2002. *Publication at Source: Scientific Communication from a Publication Web to a Data Grid*. <<http://eprints.ecs.soton.ac.uk/archive/00007852/>>.

Henry, G. 2003. On-line publishing in the 21-st Century: Challenges and Opportunities. *D-Lib Magazine*, Volume 9, Issue 10. <[doi:10.1045/october2003-henry](https://doi.org/10.1045/october2003-henry)>.

Lynch, C. 2003. Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL Bimonthly Report* 226. February 2003, <<http://www.arl.org/newsltr/226/ir.html>>.

Payette, S., and Staples, T. 2002. The Mellon Fedora Project: Digital Library Architecture Meets XML and Web Services. *European Conference on Research and Advanced Technology for Digital Libraries*, Rome, Italy, September 2002. <<http://www.fedora.info/documents/ecdl2002final.pdf>>.

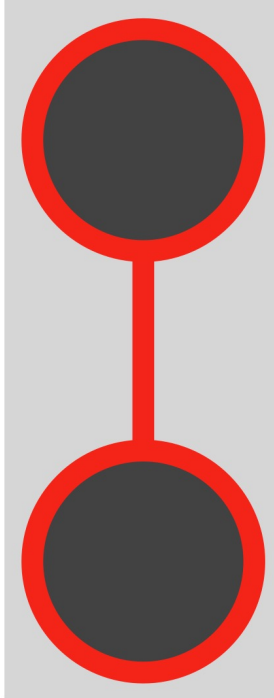
Pöschl, U. 2004. Interactive Journal Concept for Improved Scientific Publishing and Quality Assurance. *Learned Information*, Volume 17, Number 2, pp 105-113. <[doi:10.1087/095315104322958481](https://doi.org/10.1087/095315104322958481)>.

Reich, V. and Rosenthal, D. 2001. LOCKSS: A Permanent Web Publishing and Access System. *D-Lib Magazine*, Volume 7, Issue 6. <[doi:10.1045/june2001-reich](https://doi.org/10.1045/june2001-reich)>.

Roosendaal, H., and Geurts, P. 1997. Forces and functions in scientific communication: an analysis of their interplay. *Cooperative Research Information Systems in Physics*, August 31—September 4 1997, Oldenburg, Germany. <<http://www.physik.uni-oldenburg.de/conferences/crisp97/roosendaal.html>>.

OK, so what are we going
to do about it?

Robust Links



<https://robustlinks.mementoweb.org/>
<https://robustlinks.mementoweb.org/about/>
<https://robustlinks.mementoweb.org/spec/>

“Robustifying Links To Combat Reference Rot”
<https://journal.code4lib.org/articles/15509>

Robust Links

1. **Create a snapshot** of referenced resources in a public web archive



<http://web.archive.org/>

archive.is
webpage capture

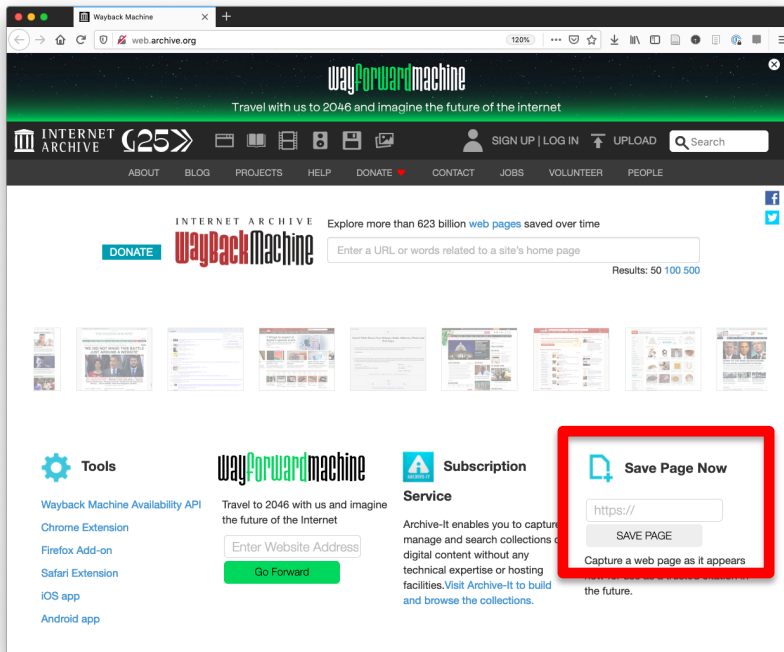
<https://archive.md/>

perma.cc ∞

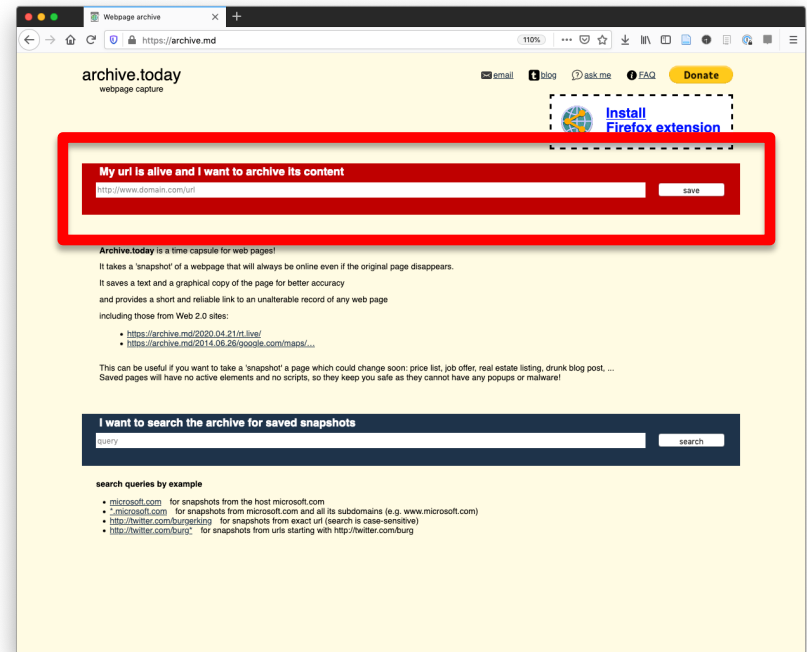
<https://perma.cc/>

Robust Links

1. Create a snapshot of referenced resources in a public web archive



<http://web.archive.org/>



<https://archive.md/>

When to create an archival copy?

- As close as possible to the time of linking!
 - Aka time of writing
 - By the author
- Or at the time of:
 - Submission
 - conference or journal submission systems
 - Publication
 - commercial publishers, pre-prints
 - Aggregation, “Consumption”
 - 3rd party such as CORE, institutional repositories

Common Practice

Capture datetime

^ Heyerdahl 1961 However, Alfred Metraux pointed out that the rubble filled Rapanui walls were a fundamentally different design to those of the Inca, as these are trapezoidal in shape as opposed to the perfectly fitted rectangular stones of the Inca. See also [this FAQ](#) at the [Wayback Machine](#) (archived 11 October 2007)

URL of archived snapshot

<https://web.archive.org/web/20071011083729/http://islandheritage.org/faq.html>

- Link to (ideally) immutable memento
- Original URL “visible”

But what if...

URL of created snapshot is:

- <http://archive.is/MTMKu>
 - <https://perma.cc/9265-T4NB>
 - <https://www.mummify.it/XbmcMfE3>
-
- It is impossible to visit the original URI, if desired
 - It requires the permanent existence/uptime of the archive that holds the snapshot
- One link rot problem replaced by another**

Robust Links

1. Create a snapshot of referenced resources in a publically available web archive
2. **Robustify links** with:
 - URL of archived snapshot (memento)
 - datetime of archiving/linking
 - original URL of resource

Benefits:

- Can visit live version of referenced resource
- Can visit created memento of referenced resource
- Original URI + capture datetime allows finding an appropriate capture in all web archives
- Uniform, machine-actionable

HTML Link

```
<a href="http://www.whitehouse.gov"  
The White House</a>
```

<http://robustlinks.mementoweb.org/spec>

Robustified Link

```
<a href="http://www.whitehouse.gov"  
data-versionurl="https://perma.cc/9265-T4NB"  
data-versiondate="2015-11-18">
```

```
The White House</a>
```

<http://robustlinks.mementoweb.org/spec>

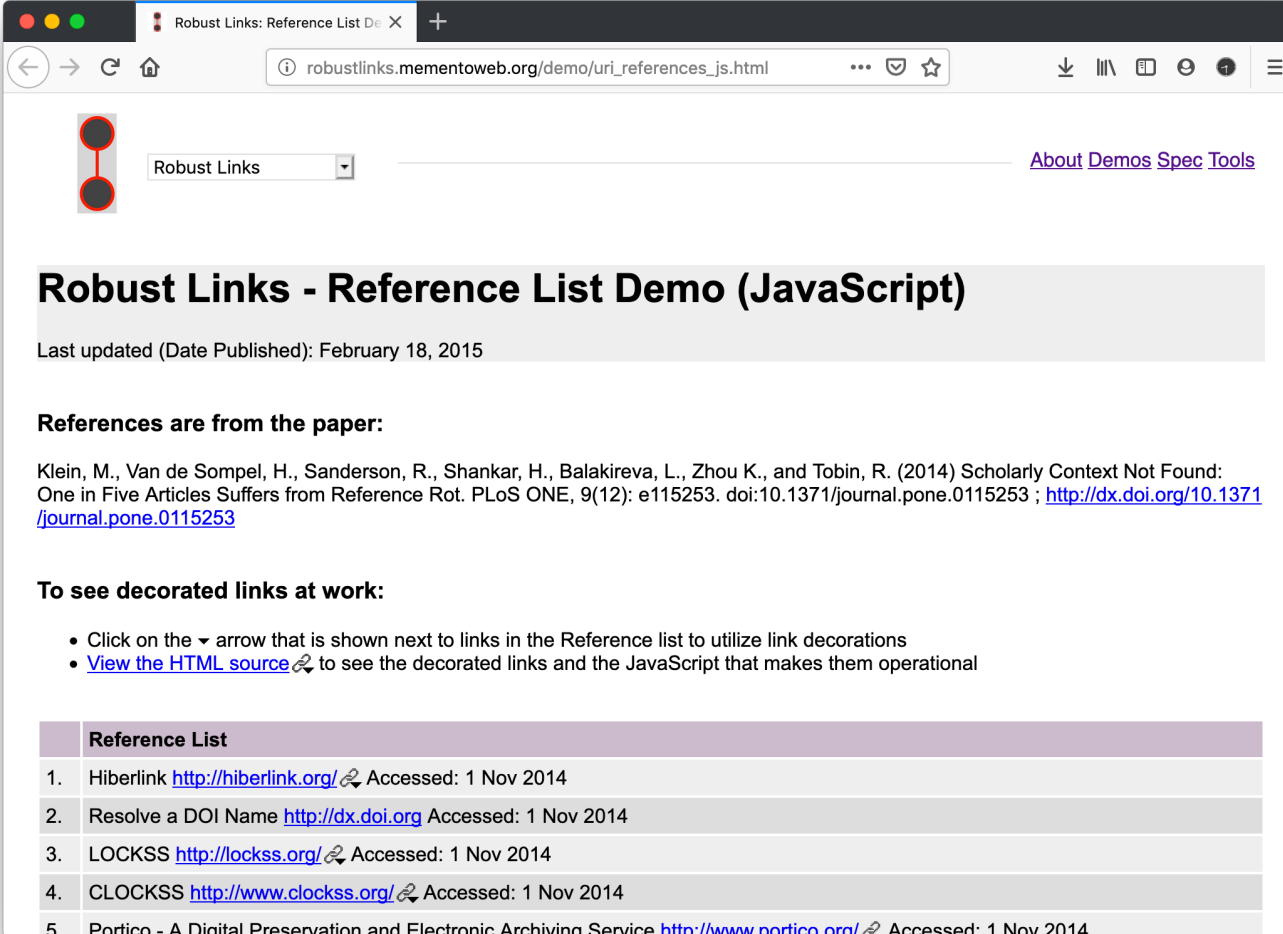
Robustified Links

```
<a href="https://perma.cc/9265-T4NB"  
data-originalurl="http://www.whitehouse.gov"  
data-versiondate="2015-11-18">
```

The White House

<http://robustlinks.mementoweb.org/spec>

Link Decoration in Action



Robust Links: Reference List Demo

robustlinks.mementoweb.org/demo/uri_references_js.html

Robust Links

About Demos Spec Tools

Robust Links - Reference List Demo (JavaScript)

Last updated (Date Published): February 18, 2015

References are from the paper:

Klein, M., Van de Sompel, H., Sanderson, R., Shankar, H., Balakireva, L., Zhou K., and Tobin, R. (2014) Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot. PLoS ONE, 9(12): e115253. doi:10.1371/journal.pone.0115253 ; <http://dx.doi.org/10.1371/journal.pone.0115253>

To see decorated links at work:

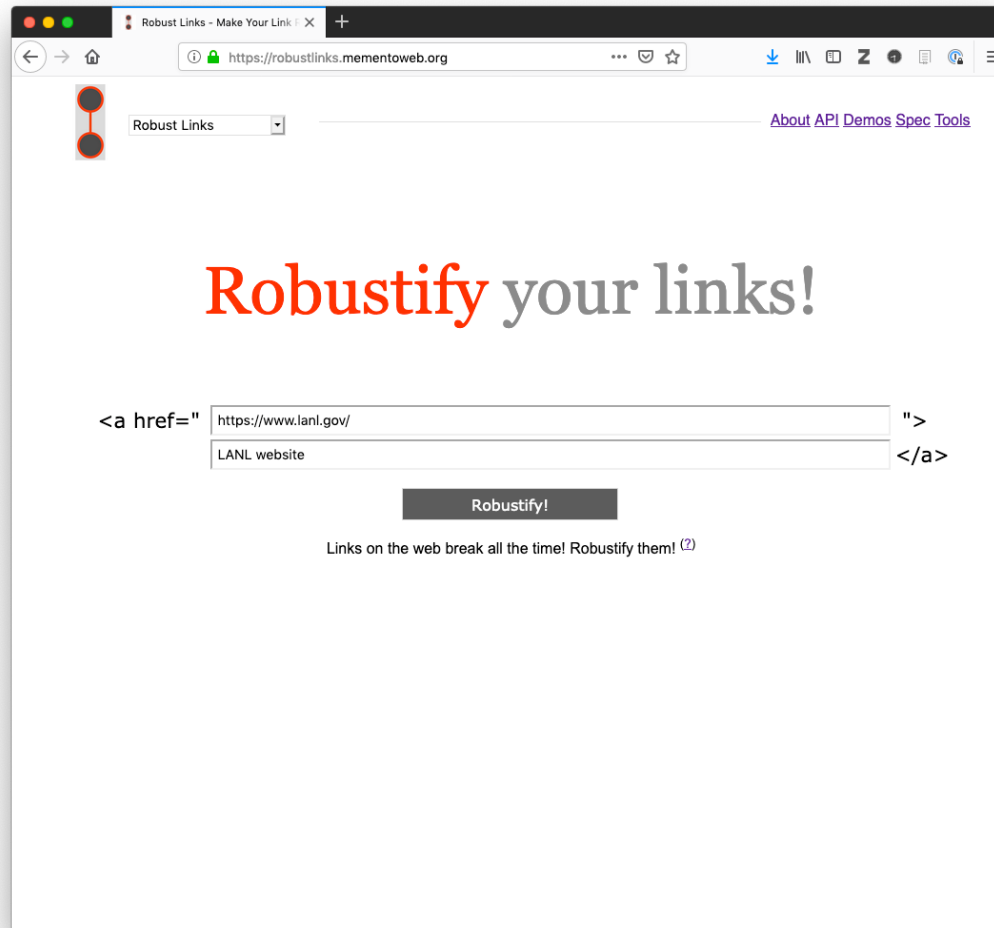
- Click on the ▼ arrow that is shown next to links in the Reference list to utilize link decorations
- [View the HTML source](#) to see the decorated links and the JavaScript that makes them operational

	Reference List
1.	Hiberlink http://hiberlink.org/ Accessed: 1 Nov 2014
2.	Resolve a DOI Name http://dx.doi.org Accessed: 1 Nov 2014
3.	LOCKSS http://lockss.org/ Accessed: 1 Nov 2014
4.	CLOCKSS http://www.clockss.org/ Accessed: 1 Nov 2014
5.	Portico - A Digital Preservation and Electronic Archiving Service http://www.portico.org/ Accessed: 1 Nov 2014

http://robustlinks.mementoweb.org/demo/uri_references_js.html

Help creating Robust Links?

<https://robustlinks.mementoweb.org/>



The screenshot shows a web browser window with the address bar displaying `https://robustlinks.mementoweb.org`. The page has a header with a logo on the left, a dropdown menu labeled "Robust Links", and a link to "About API Demos Spec Tools" on the right. The main heading is "Robustify your links!". Below this is a form for creating a robust link. The form consists of two input fields: the first is labeled "`, and to the right of the second field is ``. Below the form is a button labeled "Robustify!". At the bottom of the page, there is a text line: "Links on the web break all the time! Robustify them! ⁽²⁾".

<https://robustlinks.mementoweb.org/>

The screenshot shows a web browser window with the address bar displaying `https://robustlinks.mementoweb.org/robustify/?anchor_t...`. The page has a header with a "Robust Links" dropdown menu and links for "About", "API", "Demos", "Spec", and "Tools". The main content area is titled "Your Robust Link" and contains two steps for generating a robust link.

Step 1: Copy one of the below HTML snippets and paste it into your web page

Choose this snippet if you want the link text to lead to the live web resource `<https://www.lanl.gov/>`. Your Robust Link will appear as `LANL website` in your own web page.

```
<a href="https://www.lanl.gov/" data-versionurl="https://archive.li/wip/FvPQD" data-versiondate="2020-08-02">LANL website</a>
```

— OR —

Choose this snippet if you want the link text to lead to the memento `<https://archive.li/wip/FvPQD>` in a web archive. Your Robust Link will appear as `LANL website` in your own web page.

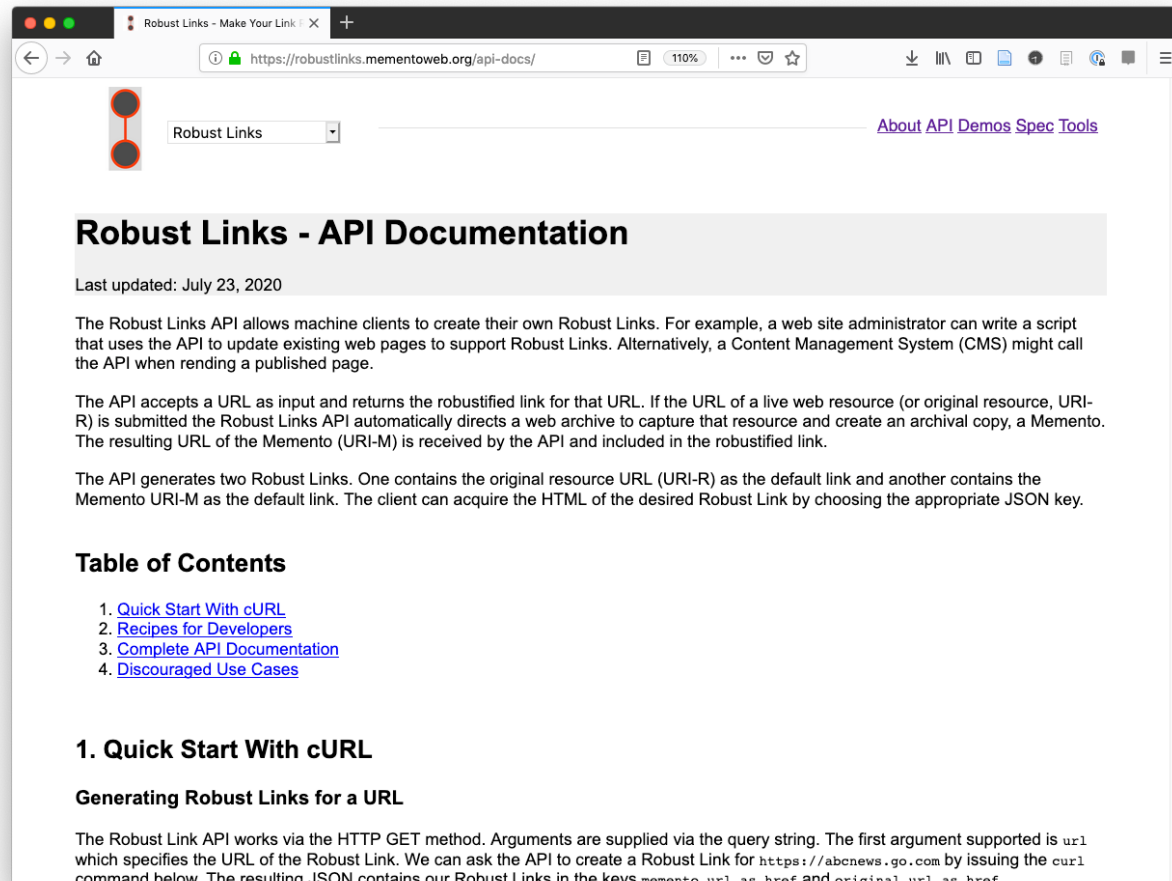
```
<a href="https://archive.li/wip/FvPQD" data-originalurl="https://www.lanl.gov/" data-versiondate="2020-08-02">LANL website</a>
```

Step 2: Copy this HTML snippet and paste it inside the HEAD tag of your web page

```
<!-- RobustLinks CSS -->
<link rel="stylesheet" type="text/css" href="https://doi.org/10.25776/z58z-r575" />
<!-- RobustLinks Javascript -->
<script type="text/javascript" src="https://doi.org/10.25776/h1fa-7a28"></script>
```

At the bottom of the page is a button labeled "Robustify Another Link".

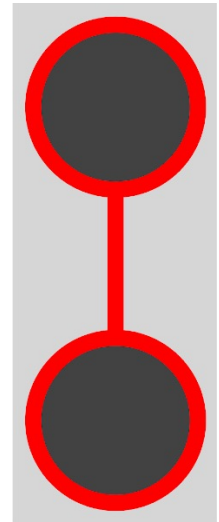
<https://robustlinks.mementoweb.org/api-docs/>



Zotero Robust Links Extension



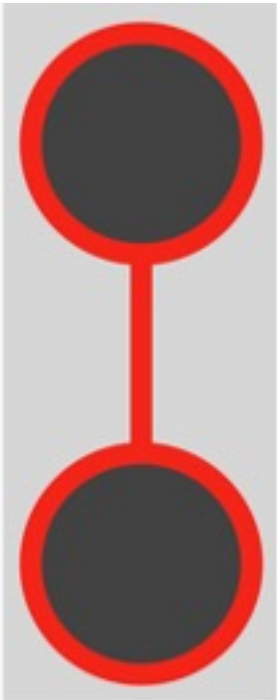
zotero +



<https://robustlinks.mementoweb.org/zotero/>
<https://github.com/lanl/Zotero-Robust-Links-Extension>

Robustify Your Links!

For better stewardship
of references to web resources
in digital scholarship



Martin Klein

mklein@lanl.gov

[@mart1nkle1n](https://twitter.com/mart1nkle1n)

<http://orcid.org/0000-0003-0130-2097>

Research Library
Los Alamos National Laboratory

```
<a href="http://www.whitehouse.gov"  
  
data-versionurl="https://perma.cc/9265-T4NB"  
data-versiondate="2015-11-18">  
  
The White House</a>
```